DEFENCE AND CIVIL INSTITUTE
OF ENVIRONMENTAL MEDICINE

Canada

March 1986 DCIEM No. 86-R-44

A COMPARATIVE EVALUATION OF THE TRACOMETER AND THE CRITICAL TRACKING TASK AS TESTS OF ALCOHOL INTOXICATION

Y. Ian Noy

Defence and Civil Institute of Environmental Medicine
1133 Sheppard Ave West
Downsview, Ontario
M3M 3B9

TABLE OF CONTENTS

ABSTRACT 3

INTRODUCTION 4

BACKGROUND 5

METHOD 8

Experimental Design 8
Subjects 9
Equipment 10
Procedures 13

RESULTS 16

Learning 16
Baseline 18
Dose-Response Relationships 18
Discriminability 22

DISCUSSION AND CONCLUSION 27

ACKNOWLEDGEMENTS 29

REFERENCES 30

APPENDIX A: Dose-Response Curves for Tracometer Variables

ABSTRACT

An experiment was conducted to investigate the feasibility of developing an in-vehicle performance-based drunk driver detection and deterrence system. The experiment compared two well-developed psychomotor test devices, the Tracometer and the Critical Tracking Task (CTT), for their ability to detect alcohol intoxication. The Tracometer employed a pursuit tracking task whereas the CTT employed a compensatory tracking task.

Twenty male subjects, aged 19 - 30 were trained on the Tracometer and the CTT over a three day period. On two subsequent test days, they were treated with either placebo or alcoholic beverages designed to bring their peak Blood Alcohol Concentration (BAC) to 1200 mg/l. On each day, the subjects were tested once before treatment and six times after treatment at hourly intervals as their BAC levels declined.

The results indicate that the Tracometer was more sensitive than the CTT to the the effects of alcohol over a wide range of BAC levels. At a BAC level of 800 mg/l, performance on the Tracometer deteriorated by 16% whereas performance on the CTT deteriorated by only 7%. Using optimized pass/fail protocols, the Tracometer failed 60% of subjects with BAC in the range 800-1000 mg/l, whereas the CTT failed 34%. At BAC levels above 1000 mg/l, the Tracometer failed about 85% whereas the CTT failed 40%. These results clearly illustrate the potential of the Tracometer as a test of alcohol intoxication. The continued development of the Tracometer as an effective drunk driver countermeasure is recommended.

INTRODUCTION

Drunk driving is perhaps the single most serious factor contributing to the incidence of traffic accidents. Recent studies performed by the Traffic Injury Research Foundation (Beirness et al., 1985) found that approximately 60% of driver fatalities tested positive for blood alcohol, 80% of whom had BAC levels in excess of the statutory limit for driving. While there may be some controversy over exact figures, it is estimated that between 1000 to 2000 deaths in Canada are caused by drunk drivers.

The alcohol-crash problem arises from the interaction of highly complex socio-economic forces within our society. A variety of countermeasures have been initiated over the years in efforts to reduce drunk driving exposure, including punitive measures, programs designed to bring about a change in social attitudes towards alcohol abuse, information campaigns, economic measures targeted at the alcohol industry and consumers, and a variety of legislative measures designed to reduce the consumption of alcohol. Nevertheless, the alcohol-crash problem has proven to be remarkably persistent.

One direct approach to preventing intoxicated persons from driving is to require them to pass a breathalyzer test prior to operating the vehicle. At the present time, however, breathalyzer technology precludes their use in vehicles. A somewhat less direct approach is to test for intoxication using tests of skilled performance. A variety of behavioural tests have been investigated over the past two decades in efforts to develop a device which can reliably detect drunk drivers (Snyder 1984). Conceptually, these devices comprise two fundamental components; (a) a detection component, and (b) a deterrence component. Performance decrements on the behavioural tests are intended to be used for predicting the driver's BAC level and are not necessarily regarded as direct evidence of impairment of driving ability. Such devices are still in the experimental stage and although none has fully met all of the stringent criteria required of such a device, progress has been encouraging. For a more complete review of the history of this area of research the reader is referred to Snyder (1984).

The present research examines two behavioural test devices, the Tracometer and the Critical Tracking Task (CTT), for their potential for detecting drunk drivers having BAC levels near the Canadian legal limit. Previous research has shown both to be sensitive to the effects of alcohol. These devices were selected for their reported sensitivity to alcohol intoxication and because they were fully engineered and available.

It should be emphasized that the devices were evaluated in their present form. There was no attempt to modify either device with respect to hardware or software. It is recognized that some opportunity may exist to improve the sensitivity of the devices through design modifications.

BACKGROUND

Research sponsored by the US Department of Transport National Highway Traffic Safety Administration (NHTSA) has led to the development of the Drunk Driving Warning System (DDWS), a vehicle-mounted system which requires the driver to pass a brief test using the steering wheel before the car can be driven in the normal manner. Theoretically the DDWS can employ any test as its central component, and at least four different tests have undergone extensive laboratory testing. The test selected by NHTSA for use in the DDWS to be fabricated and field tested was the Critical Tracking Task (CTT), a compensatory tracking task with an unstable dynamic response, which has been shown to be sensitive to the effects of alcohol (Allen, 1983a). Although at least one test, the Divided Attention Task, was shown to have as good or better discriminability, the CTT was among the best performers and was fully engineered.

In the current version of the DDWS, The CTT display unit (a dial instrument) is installed in the vehicle adjacent to the steering column. The driver is required to pass the test prior to operating the vehicle. Using the steering wheel, the driver is required to keep the pointer of the instrument in the centre of the display against displacements introduced by a computer. As the trial progresses the displacements increase in magnitude until it becomes impossible to maintain control of the pointer. The trial ends when the driver loses control of the pointer, eg., the pointer reaches either extreme positions of the display. The level of instability, lambda, at which this occurs is computed as the trial score. (Lambda is a gain in the algorithm driving the pointer which serves to maintain a constant relationship between pointer position and tracking error. That is, the greater the error, the greater the forcing function. During the trial, lambda is gradually increased until critical instability occurs). The lambda at which critical instability occurs is inversely related to the driver's 'effective delay time'.

In the DDWS, the computer compares each test score with the individual's normal sober performance, determined for each driver after a period of training with the system. Pass/Fail determinations are made on the basis of whether test performance has deteriorated sufficiently below the driver's normal level. In the DDWS, a driver has four chances to pass the test. If all four attempts are failed then the driver must wait 10 minutes before he/she can retake the test. The car will operate normally only if the test is taken and passed. Otherwise, as the driver begins to operate the vehicle, the hazard lights flash, and above a speed of 10 mph, the horn honks once per second.

The DDWS was field tested in California to assess its deterrence value for persons found to be habitual Driving-While-Impaired (DWI) offenders. The results indicated that the device had some potential for deterring drunk driving, that it was accepted by law enforcement and judicial

agencies as well as the offenders, and it had received positive public and media response. However, one major difficulty was that drunk drivers could repeatedly take the test until they passed, indicative of low test-retest reliability.

Previous studies of the sensitivity of the CTT to alcohol intoxication indicated that the test could reliably detect 75-80% of drivers with BAC levels of 1500 mg/l (Allen et al. 1983a). However, at BAC levels of 1000 mg/l the detection rate was only 35-40%. Hence, more sensitive tests would appear to be required to detect drunk drivers with BACs closer to the legal limit for driving in Canada (BAC = 800 mg/l).

The relative insensitivity of the CTT at BAC levels below 1000 mg/l is a major disadvantage of this test. Two other tests have been reported to be more sensitive than the CTT to the effects of alcohol and warrant further research. These are; the Divided Attention Task (DAT) (Snyder, 1984) and the tracking task embodied in the Tracometer (Noy, 1985). One version of the DAT, the DAT-2, consisted of a "central" compensatory tracking task and a "peripheral" choice reaction time task. However, this device was not available for study at this time.

The present research, therefore, investigated the potential of the Tracometer task as an in-vehicle test of intoxication. The Tracometer was developed at the National Research Council to investigate factors affecting human cognitive and psychomotor performance. The task is a step pursuit tracking task in which the subject is required to reposition a pointer 100 times on targets which appear in random sequence at five possible locations within the display. Several measures of performance are provided. Certain of these have been shown to be sensitive to the effects of a variety of drugs and other stressors (Buck 1981). The Tracometer test differs from the CTT primarily in that it presents a pursuit tracking task rather than a compensatory tracking task.

In an earlier study (Noy 1985), the Tracometer, was compared to the CTT in the search for a more reliable test of alcohol intoxication. In that study, 20 young male volunteers were given four dose levels of alcohol in a within-subject design. Performance measures on the two tests were taken before and after alcohol administration. The results indicated that the Tracometer was more sensitive than the CTT to alcohol intoxication. The Tracometer was able to detect over 90% of legally drunk individuals (BAC > 800 mg/l) whereas the CTT was only able to detect about 40%. However, certain methodological problems were noted. For example, the two tests could not be matched with respect to presentation time and pass/fail strategies. Another problem was that subjects did not receive sufficient practice to allow their baseline performance to asymptote. Continued research was recommended to further investigate questions related to test repeatability, test presentation time and the extent of

training required to reach asymptotic performance.

The purpose of the present experiment was to verify previous findings and to examine some further issues. In particular, the objectives of the study were:

1. To obtain dose-response data over a wider range of BAC's after more extensive training.

2. To assess the reliability of dose-response relationships.

3. To investigate whether other Tracometer measures can be used which permit presentation time to be reduced.

METHOD

Experimental Design

A mixed design with one between factor (Drug treatment) and two within factors (Day and Session) was used in the experiment. A block diagram of the design of the experiment is shown in Figure 1.

FIGURE 1: EXPERIMENTAL DESIGN

		DAY 1	DAY 2
		1 2 3 4 5 6 7	1 2 3 4 5 6 7
Placebo	S1 . S10	Placebo	Alcohol (peak BAC 1200 mg/l)
Alcohol	S11 . S20	Alcohol (peak BAC 1200 mg/l)	Alcohol (peak BAC 1200 mg/l)

Twenty subjects were divided into two groups of 10; a Placebo group, and an Alcohol group. Each subject was tested on two different days. On their first test day, Day 1, subjects in the Alcohol group were treated with alcoholic beverages (individually prepared to produce an instantaneous peak BAC level of 1200 mg/l at 90 minutes after the start of drinking) whereas, subjects in the placebo group received beverages with no alcohol content. Performance measures on the two tests (Tracometer and CTT) were obtained prior to dosing, at peak BAC, and five more times at hourly intervals during the alcohol elimination phase. Hence, seven test sessions were completed by each subject on each day. On their second test day, Day 2, all subjects were treated with alcohol to a peak BAC of 1200 mg/l and retested in the same way.

Strictly speaking, this design permits a comparison of two different ways of quantifying the effects of decreasing levels of alcohol intoxication. In this experiment subjects were dosed to the same initial levels. Performance measures and corresponding blood alcohol levels were obtained in a within-subjects paradigm as BAC levels dropped. The more usual approach for obtaining dose-response data is to dose subjects to different levels of BAC and obtain performance data at the peak levels in a between-subject paradigm. The advantage of the repeated measures approach is increased efficiency since each subject provides six sets of data rather than one.

The disadvantage with the repeated measures approach is that it does not control for the effects of acute tolerance, carry-over effects or other time-dependent effects. Acute tolerance (eg., an increasing tolerance to the effects of alcohol with time since drinking), however, is an organismic phenomenon (Smart et al. 1983) which should not affect the relative sensitivity of the tests to impairment due to alcohol. Similarly, if carry-over effects were present, they would be expected to affect the organism's response to alcohol. They would not be expected to produce differential effects for the two tests. Hence, while there are limitations in the design, these were justified on the basis of efficiency (alternate designs are considerably more complicated and lengthy) and because they do not jeopardize the comparability of the tests.

The design permits comparisons between the Placebo and Alcohol groups on Day 1 to establish the magnitude of the alcohol effect. A comparison of Day 1 and Day 2 data for the Alcohol group would provide a measure of test-retest reliability. The Placebo group was treated with alcohol on Day 2 in anticipation that the Day 2 data could be pooled across all subjects, hence yielding a larger population from which dose response relationships could be established and for comparing the abilities of the two tests to discriminate between legally-drunk and legally-sober individuals.

Subjects

Twenty young male subjects between the ages of 19-30 years were selected from respondents to newspaper advertisements. They were screened on the basis of a drug use questionnaire to eliminate regular users of psychoactive drugs and to ensure that they were moderate drinkers as defined by the Cahalan, Cisin and Crossley quantity-frequency scale of alcohol use (Cahalan et al. 1969). Subjects were rated as moderate-light (drinking category 1), moderate (drinking category 2) and moderate-heavy (drinking category 3) drinkers. These ratings, based on the average amount of alcohol consumed per week, were determined using the procedure described by Cahalan et al. (1969). The number of subjects in categories 1, 2, and 3 were 5, 3 and 12, respectively.

Subjects eligible to participate in the study were required to undergo a medical examination to ensure that they were medically fit and not taking any medication. During the medical examination, each subject's weight and skinfolds were measured. Skinfold measurements were taken at the triceps, the sub-scapular and supra-iliac sites. Subjects who were accepted for the study were asked to sign a consent form to indicate that they understood the purpose and requirements of the study and were willing to participate. The skinfold measurements were needed for calculations of the subject's fat composition which was used in the preparation of the individual alcohol doses.

During the study, subjects were required to abstain from consuming the following substances: marijuana for a period of one week prior to testing; caffeine and food for 4 hours prior to testing; and, tobacco for 1-2 hours prior to testing. Subjects were allowed to smoke a maximum of 2 cigarettes during the experiment if they reported a strong need. They were paid $195.00 for participating in the study.

Equipment

(i) Breathalyzer

An Alcometer, model AE-D1, was used to measure subjects' breath alcohol concentration. The digital readout provided BAC readings within an accuracy of + or - 5 mg/l. Approximately 20 seconds were required to administer the breath test.

(ii) CTT

The CTT was part of a DDWS installed in a 1978 Chevrolet Nova which was provided by the NHTSA. It comprised four basic components: the CTT display unit and steering sensor, an electronics module including microprocessor, a seat-weight sensor, and a data logger. A more complete description of the DDWS/CTT is contained in the report by Allen (1983). In the present experiment, only the CTT module was required. Figure 2 is a photograph of the CTT display installed in the DDWS. The pass/fail criteria was set to its maximum value of 9.9 and the pass and fail indicator lights on the display unit were covered over to prevent the subject from receiving feedback. Performance on the test was displayed on a readout module located at the experimenter's station.

Typical scores ranged between 3.0 and 6.0 with higher scores indicating better performances. An average run took about 30 seconds to complete.

(iii) Tracometer

The Tracometer was on loan from the NRC. It comprised a tracking unit and control panel. Figure 3 is a photograph of the Tracometer. The task itself was a step-input, subject-paced, pursuit tracking task. The subject was seated in front of the tracking unit which consisted of five target positions (small lights). At the start of a trial one of these targets was illuminated. The subject's task was to align the pursuit pointer with the target using a control wheel. The control wheel was connected in such a way that it had to be turned in a direction opposite to the desired pointer movement. Once the pointer was accurately aligned with the target for a period of 200 msec., the target was extinguished and a new target appeared at random at one of the other four positions. A single trial consisted of 100 consecutive repetitions, or movements, of this task, divided into five segments of 20 movements each. Each segment

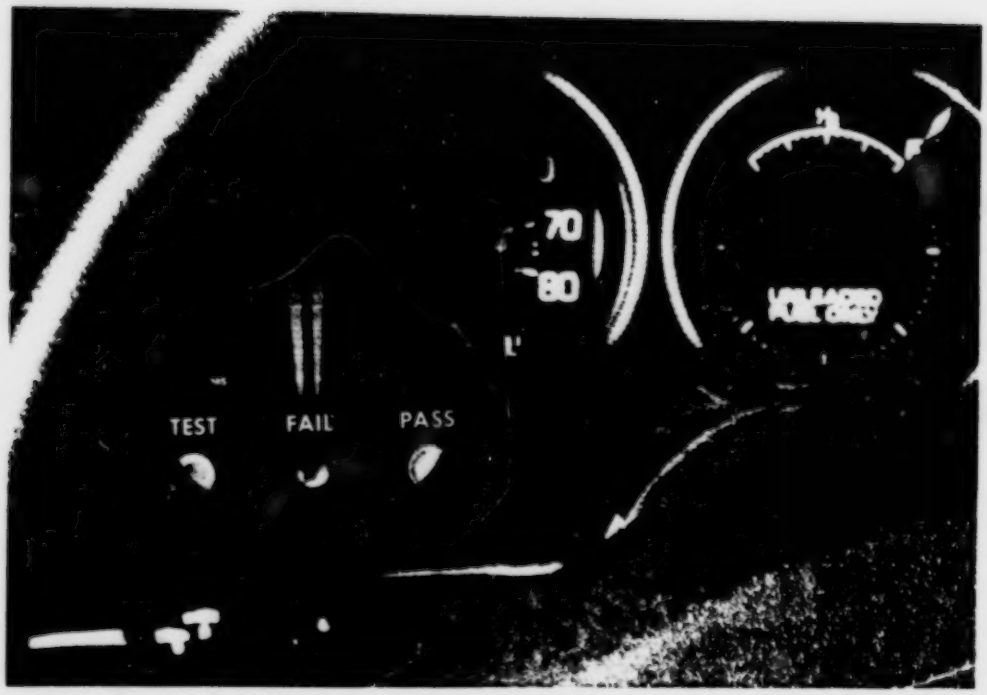


FIGURE 2: CTT DISPLAY

FIGURE 3: TRACOMETER

presented all possible movements between targets, though not in the same order. The subject was instructed to complete the 100 movements as fast as possible.

The Tracometer automatically recorded subjects' responses on a cassette tape. The data from the tapes were later reduced at the NRC, giving six categories of measures for each trial; Correct Response Time (CRT), Non-Overshoot Movement Time (NOM), Overshoot Movement Time (OM), Overshoot Score (OS), Error Score (ES) and Total Response Time (TRT). For each category, several measures were provided depending on the type of movement involved (level of variable). In addition to these measures, the experimenter manually recorded the elapsed time for each segment of 20 movements during each Tracometer trial. These data were used to derive the 5 Response Time (RT) scores of the trial. In total, there were 26 different response measures (scores) for each trial. Table 1 provides a list of the Tracometer categories and the dimensions underlying the structure of the data. For a more detailed description of the Tracometer see Buck et al. (1981, 1982, and 1983).

TABLE 1: TRACOMETER MEASURES

CATEGORY	DESCRIPTION	MEASURES	DIMENSION
Correct Response Time (CRT)	Time to initiation of movement in correct direction	4	Prob. of location of next target (0, .25, .5, .75, 1)
Non-Overshoot Movement Time (NOM)	Time per movement when no overshoot occurred	4	Distance to next target
Overshoot Movement Time (OM)	Time per movement when target was overshot	4	Distance to next target
Overshoot Score (OS)	Percentage of overshoots	4	Distance of next target to edge of display
Error Score (ES)	Percentage of movements started in the wrong direction	4	Prob. of location of next target
Response Time (RT)	Elapsed time for a segment of 20 movements	5	Segment no.
Total Response Time (TRT)	Elapsed time for the trial	1	
TOTAL		26	

Procedures

(i) Training

Each subject received two hours of training on three consecutive days, for a total of six hours. On each training day, the subject completed 100 runs on the CTT and 12 runs on the Tracometer. Practice was given in four blocks of 25 trials on the CTT interspersed with 3 blocks of 4 runs on the Tracometer. Therefore, in total, each subject received 300 training runs on the CTT (representing a total of 150 minutes of training) and 36 training runs on the Tracometer (representing a total of 90 minutes). The amount of training provided was consistent with reported acquisition times to reach asymptotic levels. Allen (1983) recommended 300 training trials of the CTT over three days. Buck (1981) reported that 16 trials on the Tracometer is sufficient. During training only, subjects were given feedback and encouraged to improve their performance.

(ii) Testing

Four subjects were scheduled for testing on any given day. Upon arrival at the laboratory, subjects provided a urine sample which they were told would be screened for the presence of drugs. In fact, the samples were not analyzed. The daily experimental schedule is depicted in Figure 4. On each day, four subjects began the experiment fifteen minutes apart.

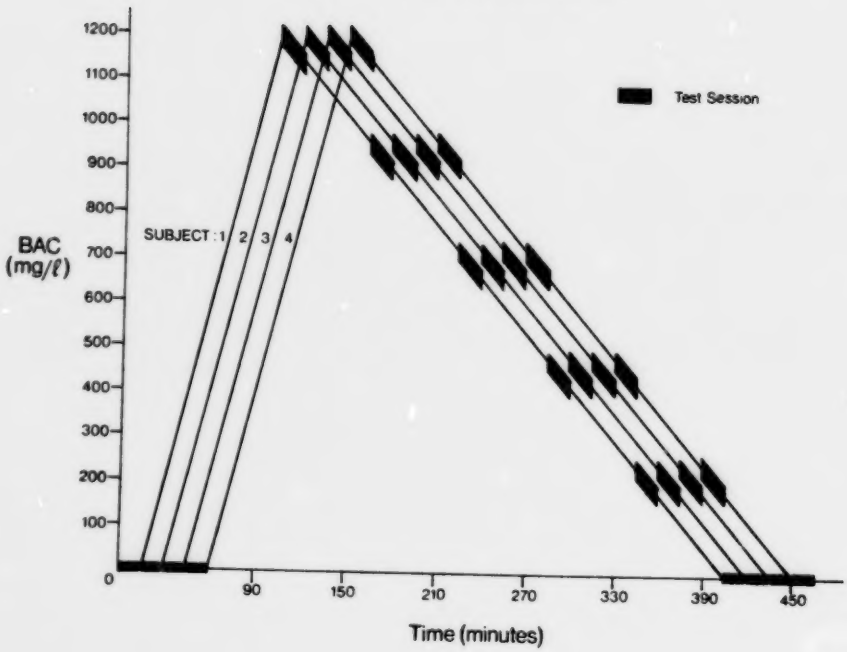
During the course of the day, each subject completed seven test sessions, each taking approximately 15 minutes to complete. Each session consisted of a breathalyzer reading, two runs on the Tracometer, and 12 runs on the CTT. As a run on the Tracometer took approximately four times as long as one on the CTT, test exposure was roughly equivalent. Moreover, each trial on the Tracometer was divided into a set of five independent and equivalent segments, each of which was roughly the length of one trial on the CTT. The first session was taken before the drug treatment to establish pre-treatment performance levels.

After the first session, the subject was taken to an adjacent lounge for treatment. Subjects were not informed about the amount of alcohol they would receive. The drinks were prepared daily for individual subjects on the basis of the subject's lean body weight.

The alcohol treatment was administered in the form of three equal strength screwdrivers with a total alcohol content of 3.475 ml of alcohol per kilogram of lean body mass, diluted with orange juice mix in the ratio of 1:1.67. This amount of alcohol was calculated, taking into account the kinetics of distribution and elimination, to be required to produce a peak blood alcohol concentration of 1200 mg/l. The peak concentration is normally achieved (after fasting) at some 25 to 30 minutes after consumption of a given dose.

Figure 4

Test Day Schedule



The clearance of alcohol is known to be zero order, at a nominal rate of 200 mg ethanol per litre per hour, assuming that renal excretion is negligible by comparison to metabolic transformation. The placebo treatment consisted of an eyedropperful of vodka floated on the subject's three drinks. The vodka was added moments before administering the drink to the subject. In this way, the initial taste and smell of the placebo was similar to the alcoholic beverages and from casual comments it was clear that most subjects believed all drinks contained some amount of alcohol. The drinks were given to the subject over the period of 60 minutes, at twenty minute intervals.

After having received the treatment, the subject rested for 30 minutes. The peak concentration was assumed to occur at 90 minutes after the beginning of drinking. At this point, the subject completed a second test session. Five more test sessions followed at hourly intervals. Between test sessions the subjects were allowed to relax in an adjacent lounge. A light lunch was given between the third and fourth session. At the completion of the sixth test session, the subject's BAC was measured to ensure that it was below 300 mg/l and he was driven home.

RESULTS

Learning

An attempt was made in the present experiment to ensure that sufficient practice was provided during training to achieve asymptotic levels. It was not possible, however, to provide equal exposure to the two tests because of the practical constraint of having to train four subjects per day and the requirement to provide at least the minimum exposure for each test recommended in the literature. Over three days of training each subject received a total of about 150 minutes of training on the CTT (representing 300 trials) and about 90 minutes on the Tracometer (representing 36 trials).

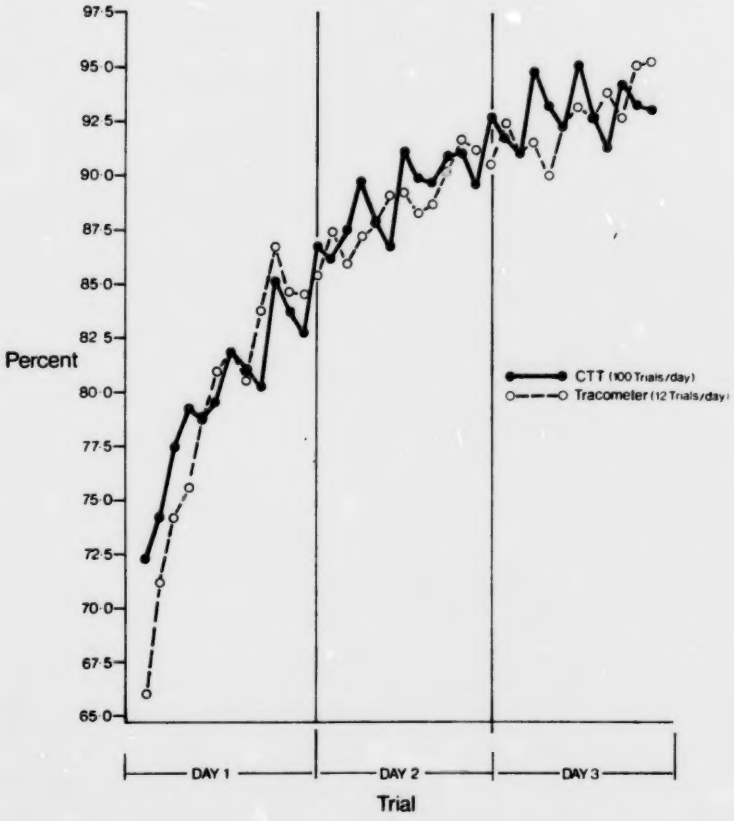
Acquisition data were recorded during training. In order to compare learning rates during the training sessions, acquisition data for both the Tracometer and the CTT were converted to a percentage of the best score attained by the subject during training. Acquisition data for both tests are shown in Figure 5. For both devices, there was a significant improvement on the first day of training, as evidenced by a highly significant main effect of Session ($p<.0001$). For the CTT, significant improvement was observed on the second day but on the third day the improvement did not reach significance. This would suggest that the subjects had reached stable performance levels on the CTT at the end of training.

For the Tracometer, significant improvement in performance was also observed on the second and the third day of training ($p<.05$), though the rate was greatly diminished on the third day. The reader should note that the number of practice runs given on the Tracometer, 36, was nearly triple the number recommended by Buck (1983). It would appear that even this number of trials may not have been sufficient to ensure that learning was complete, though performance appeared to be reaching plateau.

To investigate whether learning on the Tracometer continued into the experimental phase, a t-test was performed comparing pre-treatment performance of Day 1 with the end-of-training performance. The difference in mean performance did not reach significance at the .05 level. In addition, analysis of the placebo data did not indicate that learning continued beyond the training period. Performance data for the placebo condition were not affected by Session number. Hence, although the acquisition data indicated that performance on the Tracometer may not have stabilized during training, the results obtained in the test phase suggest that learning did not extend beyond the training phase.

Figure 5

Training Scores



Baseline

In an effort to minimize the variance due to individual differences in skilled performance, the raw scores for each subject were transformed to difference scores by subtracting the individual's baseline scores. It was necessary to derive the baseline scores from the acquisition data rather than use pre-treatment data since under the envisioned use of the test no pre-drinking scores would be available.

There were several alternative methods for calculating individual baselines from the acquisition data, all equally valid. Seven methods for the Tracometer and nine methods for the CTT were compared to determine which one(s) best predicted future pre-treatment levels. The methods evaluated are listed below.

Tracometer Baseline Measures

- o second best training score
- o second best average training session score
- o best average training session score
- o best training score
- o last training session average
- o overall average of day three
- o 95th percentile score

For the CTT, the methods were:

- o second best training session average
- o best average training session score
- o overall average of day three
- o best average half session score
- o second best average half session score
- o last training session average
- o last training half session average
- o 95th percentile score
- o NHTSA method (2nd highest block mean - .2)

The results indicated that for both tests, the 'average of the last training session' minimized the mean and variance of the pre-treatment differences scores. This method was therefore used in the calculation of all difference scores. A sensitivity analysis indicated that choice of baseline was of little significance to the results.

Dose-Response Relationships

It should be noted that although a factorial design was used to obtain the experimental data, much of the analyses used a variety of regression techniques to evaluate dose-response relationships. Instead of using Session as an independent variable in an analysis of variance, BAC level was used as the explanatory variable in regression analysis.

(i) Calculations of Actual BAC's

BAC time histories were generated for each subject using regression procedures. They confirmed the validity and accuracy of the dosing algorithm. The mean peak BAC observed for subjects that received alcohol was 1197.2 mg/l and the mean elimination rate was 200 mg/l/hr. The linear correlation coefficients for the relationship between BAC and time ranged from .969 to .999.

It will be recalled that BAC levels were measured at the start of each session. However, each session took about 15 minutes to complete. Hence, there was a slight decline in BAC within a session. (The largest possible error due to this problem was about 50 mg/l.) Therefore, measured BAC values were adjusted using individual regressions of BAC with time to correct for declining BAC levels within a session, thus achieving a high degree of correspondence between test scores and momentary BAC levels.

(ii) Pooling Day 2 Data

The Day 2 data from both groups of subjects (both placebo and alcohol groups received alcohol on Day 2) were compared using homogeneity of slopes and analysis of covariance (with BAC as the covariate and Group as the class variable). For both the CTT and the Tracometer, the difference between the two groups was found to be not statistically significant at the .05 level. Hence, it was possible to pool the data from both groups to obtain a larger dataset for examining dose-response relationships.

(iii) Selection of Best Tracometer Score

As shown in Table 1, the Tracometer provides several scores for each trial, unlike the CTT which provides only a single score. Each trial of the Tracometer yields seven different variables, each having one or more levels, for a total of 26 different scores. It was necessary, therefore, to derive one Tracometer score which is maximally sensitive to the effects of alcohol. For the analysis, the levels for each variable were collapsed, yielding seven possible variables. Stepwise regression techniques were used to identify the best multiple predictor model.

It is important to note that these analyses were simply performed to select from among the several possible scores, the one (or ones) which was most sensitive to the effects of alcohol. Neither design parameters nor the scoring methods for the Tracometer were manipulated since they were internal to the system.

The results indicated that difference scores for Overshoot Movement Time (OM), and Total Response Time (TRT) were highly sensitive to BAC levels, each having an R-square value of .45. Because of the relative simplicity in deriving the TRT score, it was selected as the Tracometer response

variable for the dose-response analyses. Including other predictors in the model did not significantly improve the fit. (Regression analyses with six variables in the model increased the R-square value to .52.)

(iv) Dose-Response Curves

Regression techniques were used to analyse dose-response relationships separately for the two tests. In order to compare dose-response curves for the two tests, ratio scores were calculated for each subject and plotted as a function of BAC. The ratio score was calculated as Raw/Baseline. The resultant linear functions are shown in Figures 6A and 6B. Other models (eg., logarithmic and quadratic) were attempted but did not result in a better fit of the data.

The the dose-response curves (Figure 6) indicate that both tests were sensitive to the effects of alcohol (eg., TRT increased with BAC and the CTT score decreased with BAC), as expected. In both cases, the slopes were significantly different from zero at the .001 level of significance. Comparison of the two curves indicate that the Tracometer was more sensitive to alcohol, as evidenced by the steeper slope. At 800 mg/l, Tracometer TRT increased by 16% and the CTT performance declined by 7%. At a BAC level of 1000 mg/l, the Tracometer scores increased 20% above the baseline levels whereas the CTT score declined by only about 9%. The decrement in performance on the CTT is consistent with findings reported in the literature (Allen et al. 1983).

It is also instructive to compare the 95% confidence limits shown in Figure 6. The narrower band for the Tracometer indicate that the Tracometer data were more consistent, or less variable, across subjects. The comparatively steeper slope and and reduced variability of the Tracometer suggest that it may be a better test of alcohol intoxication. The accuracy with which these tests can detect intoxicated subjects (at different levels of BAC) are compared in a later section, Discriminability.

(v) Test-Retest of Dose Response Curves

The Day 1 and Day 2 data from the Alcohol group were compared using homogeneity of slopes and analysis of covariance. For both the CTT and the Tracometer, the difference between the two days was found to be not statisticlly significant at the .05 level. That is, there no reliable differences were found between the dose-response curves for day 1 and day 2. Hence, test/retest reliability was confirmed.

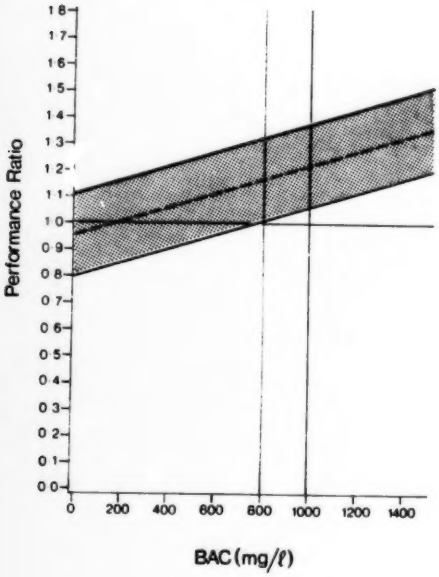
(vi) Chronic Tolerance

Individual drinking habits were used to categorize the subjects as light-moderate, moderate, and heavy-moderate drinkers in order to determine the effects of chronic tolerance.

Dose-Response Curves

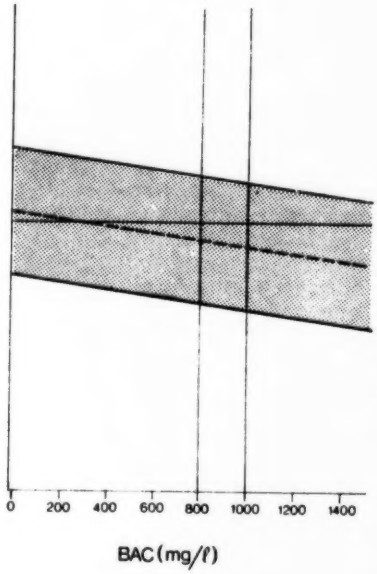
Expected Means and 95% Confidence Limits

Figure 6A



Tracometer

Figure 6B



CTT

Analysis of covariance was used to compare the dose-response curves for the three groups (BAC as covariate and Drinking Code as class variable). For the Tracometer, no reliable differences were found between groups. For the CTT, the moderate drinkers were most affected by alcohol, followed by light and heavy drinkers. For this test, the differences between the groups were significant at the .05 level. It is not clear why the CTT, but not the Tracometer, could discriminate among groups when it appears to be less sensitive to the effects of alcohol. Because of the large imbalance in the number of subjects in each group (5 in group 1, 3 in group 2 and 12 in group 3), further analysis of the differences among the groups was not possible.

These results suggest, however, that individual responses are affected to a degree by one's history of alcohol use. This may imply that the tests could produce higher proportions of misses for the heavy drinkers. However, it might also be argued that heavy drinkers are less affected by alcohol at 800 mg/l BAC and that the implications of misses for these drivers are less critical. Further research is necessary to investigate the performance patterns of heavy drinkers and determine whether the Tracometer would be as effective in detecting this group of drivers at higher BAC limits.

Discriminability

This section examines how accurately the tracking tasks discriminated between intoxicated and sober subjects. The basic measure of discriminability was the failure rates achieved by the tests. A failure (detection) was defined as a prediction by the test that the subject's BAC level was greater than 800 mg/l (the legal limit). Pass/fail decisions were computed on the basis of the magnitude of the deviation from individual baseline levels. That is, difference scores (the difference between trial score and baseline score) were compared with a population threshold, which was derived from the data. A failure occurred when the subject could not reach criterion level of performance within a given number of trials.

To compute population failure rates, it was first necessary to determine for each test the optimal test strategy and associated pass/fail criterion (threshold), jointly referred to as the test protocol. Test strategy refers to the number of trials permitted to obtain a passing score. Because data were obtained for several repeated trials at each level of BAC, it was possible to compare a variety of different test strategies. Similarly, for each test strategy, it was possible to define several different pass/fail criteria. For example, for a one pass in four test strategy, pass/fail criteria could be defined to fail 0, 1, 5, 25, or 50% of subjects within certain BAC ranges. Some reasonable criteria might be: 25% failure rate for subjects with BAC between 500 and 700 mg/l; 5% failure rate for subjects with BAC of 0 mg/l, 10% failure rate for subjects with BAC between 0 and 300 mg/l,

and so on. It is apparent that for each strategy numerous such criteria exist and that the selection of any one criterion is somewhat arbitrary.

The method of selecting the optimal protocol for each test began with the definition of a set of possible test strategies and pass/fail criteria. Then, failure rates were computed for a series of BAC ranges. The optimal test protocol (for a given test) was defined as the protocol which achieved the best combination of low sober failure rate, high intoxicated failure rate, and short testing period for sober subjects. In order to make valid comparisons between the tests, it was necessary to fix the sober failure rate (eg., the false positive rate for subjects with BAC = 0 mg/l). The sober failure rate was arbitrarily fixed at an upper limit of 5%.

Although the dose-response data presented earlier were based on the TRT as the measure of performance on the Tracometer, the discriminability analyses were based on RTs (segment scores). (Both RT and TRT provided mean response time per movement; however, RT was based on 20 movements while TRT was based on 100 movements). RTs were not part of the formal Tracometer database and hence were not used for the dose-response analyses. However, in an effort to control for the effects of test duration it was decided to use RT since it was more comparable to the CTT score (than TRT). The duration of one segment of the Tracometer trial, about 25 seconds, was roughly equivalent to the duration of a CTT trial (about 30 seconds). Regression analysis indicated that there were no statistically significant differences between RT and TRT response curves (see Appendix A for the dose-response curves).

It should be noted that the similarity between the RT and TRT scores has important implications for the viability of the Tracometer as a practical countermeasure. This finding suggests that Tracometer test duration can be reduced by a factor of 5 without much loss in test sensitivity to alcohol.

For the Tracometer, two test strategies were selected; one pass out of one segment (T1/1), and one pass out of 5 segments (T1/5). For the CTT, similar test strategies were selected; one pass out of one trial (C1/1), and one pass out of four trials (C1/4). It should be noted that T1/5 is roughly equivalent to C1/4 in terms of presentation time. Similarly, T1/1 is roughly equivalent to C1/1.

Comparisons of failure rates were determined at different ranges of BAC to determine how discriminability changes with BAC. The BAC data were divided into seven categories: Placebo (0 mg/l), Pre-treatment (0 mg/l), BAC range 0-299 mg/l, BAC range 300-499 mg/l, BAC range 500-799 mg/l, BAC range 800-999 mg/l and BAC greater than or equal to 1000 mg/l.

Table 2 provides comparative failure rates (percentages) for several 'reasonable' test protocols. As it turned out, the optimal test protocols for the two tests were not the

same, though they both had less than 5% failure rates for pre-treatment data. The optimal protocol for each test is indicated in Table 2 with an asterisk. It was noted that, for any given protocol, the Tracometer consistently failed more intoxicated subjects.

TABLE 2: Failure Rates (Percentages) Using Several Different Protocols For Strategies: T(1/5) and C(1/4)

Criteria	Test	Pre-treatment	BAC LEVEL 0-299	300-499	500-799	800-999	>1000
25% fail with BAC 500-799	T*	1.25	2.35	11.86	24.10	59.32	85.14
	C	4.17	9.49	13.79	24.80	27.96	30.61
10% fail with BAC 300-499	T	0.00	2.35	8.47	20.48	49.15	75.68
	C	1.67	6.75	9.20	16.00	20.43	19.39
5% fail with BAC 0-299	T	6.25	4.71	30.51	40.96	72.88	91.89
	C	10.83	10.95	9.63	31.63	37.63	43.88
1 S fail Pre-RX	T	1.25	1.17	16.94	32.94	57.62	75.64
	C*	3.33	2.72	4.59	14.81	34.34	40.81

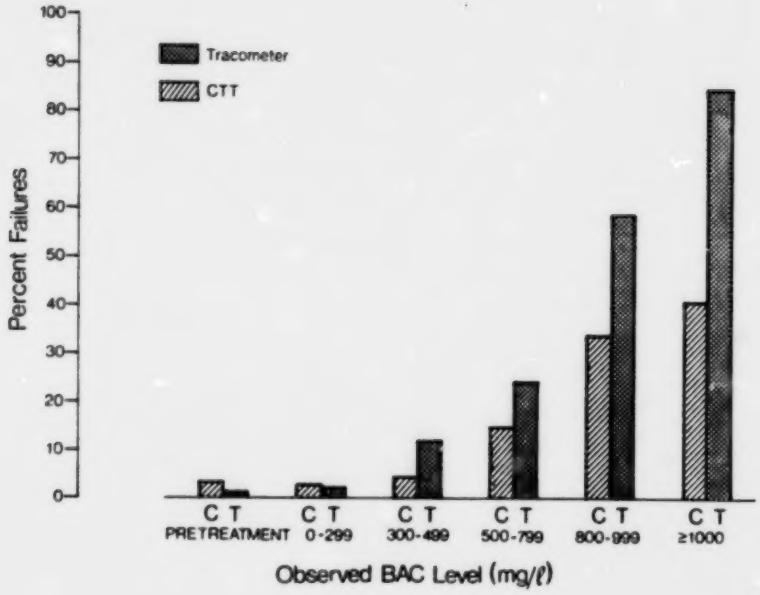
The optimal test protocol for the Tracometer was a one pass out of five strategy with a pass/fail criterion of 5.22 (difference threshold used for all subjects), designed to fail 25% of subjects with BAC levels between 500 and 799 mg/l. Analysis of other test strategies indicated that it was possible to achieve higher failure rates for intoxicated subjects without significantly increasing the sober failure rate. This, however, would have required two complete Tracometer runs, resulting in a presentation time far beyond what might be considered reasonable for the intended application.

The optimal test protocol for the CTT consisted of a one pass out of four trials strategy with a pass/fail criterion of -.1708, designed to fail one sober (0 mg/l) subject. (Note, the negative value reflects the fact that impairment results in lower scores on the CTT). This protocol failed the largest number of intoxicated subjects within the 5% failure restriction of sober subjects. Other protocols were able to fail more intoxicated subjects, but the failure rate for sober subjects would have risen above 10%.

The failure rates achieved for both tests using the optimal protocols are illustrated in Figure 7. For both tests, the sober failure rates were well below the 5% level. As BAC levels increased, the Tracometer failure rates increased more rapidly than did the CTT failure rates. At BAC levels of 800 mg/l (the Canadian legal driving limit), the Tracometer failed 60% of intoxicated subjects while the CTT failed 34%. At BAC levels of 1000-1200 mg/l, the Tracometer failed 85.14% of the subjects whereas the CTT failed 40.81%. The Tracometer appeared to have better discriminability than the CTT at BAC levels beyond 800 mg/l. However, as Figure 7 illustrates, the Tracometer also had a higher failure rate than the CTT at BAC levels between 300 and 800 mg/l. These subjects, though affected by alcohol, were legally permitted to drive. However, failures in this range should not necessarily be considered as false positives. In Ontario, for example, it is a violation of Provincial laws to drive with BAC levels above 500 mg/l. Hence, the Tracometer could potentially be useful in detecting this group of drivers as well. The true false positives were the failures at BAC levels between 1 and 500 mg/l. As Figure 7 indicates, the failure rate at this BAC range was below 10% (when combining the 0-299 and 300-499 ranges).

Figure 7

Percent Failure



DISCUSSION AND CONCLUSION

The purpose of this study was to explore the potential for developing an in-vehicle intoxication warning devices (IWD) as a means for deterring drunk driving. This study has demonstrated that tracking devices such as the Tracometer have considerable potential for detecting drunk drivers. The Tracometer in particular, was shown to be highly sensitive to the effects of alcohol and that test protocols reliably detected 60% of subjects having BAC levels of 800 mg/l and about 90% of subjects having BAC levels of 1000 mg/l or higher. These results are consistent with earlier studies of the Tracometer (Noy 1984, Buck 1981).

It was of interest to note that despite the obvious decrements in performance, the dosage of alcohol did not produce gross signs of drunkenness in the subjects.

The second test that was evaluated, the CTT, did not exhibit the same degree of sensitivity to alcohol and did not discriminate as well at intermediate and high levels of intoxication. CTT performance scores also exhibited much higher within-subject variability. The CTT test results obtained in this study compare favourably with results reported in other studies (Allen 1983a, Noy 1984). The dose-response relationship is similar in regards to the range of scores obtained and the mean performance decrement at 1000 mg/l. Furthermore, the discriminability of the CTT using the strategy developed in the U.S. yielded very similar failure rates - about 40% at the 1000 mg/l BAC level.

A major difference between the two tracking tests evaluated in this study was that the Tracometer presented a pursuit tracking task whereas the CTT presented a compensatory tracking task. It is not clear, however, whether the differences in the results are due to the specific cognitive processes which underlie compensatory and pursuit tracking or whether they are due to the specific implementation of the two tasks. In this respect, the results of this study should not be generalized beyond the specific devices tested.

The dose-response relationships for the Tracometer can be adequately described using linear models. Test-retest reliability of this model has been demonstrated. A variety of Tracometer measures were evaluated. The most sensitive and reliable was the Total Response Score (the time to complete 100 movements). However, it was observed that Response Time (time to complete 20 movements) exhibited the same degree of sensitivity and discriminability as did the Total Response Time Score. This finding was highly significant since it implied that the length of each Tracometer trial could be reduced by 80% without a significant loss in sensitivity or discriminability. Response Time can be obtained in about 30 seconds, a presentation time which is comparable to the CTT and which is not unreasonable. The failure rates shown in Figure 7 were based on Response Time data for 20-movement

segments.

It should be recognized that for the Tracometer test there was some latitude for adjusting the pass/fail criterion to obtain lower sober failure rates without adversely affecting the effectiveness of the test. This opportunity did not exist for the CTT test since the failure rate at BAC levels above 800 mg/l (40%) was already too low.

The possible implementation of the IWD as a drunk driving countermeasure can take many forms. It can be specifically targeted for use by convicted DWI offenders, as was the case in the California study. It can be mandated for use on public transportation or commercial vehicles. It can also be offered as optional or after-market equipment to companies owning a large fleet of vehicles or to the general public. However, it should be recognized that drivers using these systems would require some initial training to establish individual criteria.

Whether and how such devices may be used in the future will depend largely upon the reactions of groups representing business, labour, the public, and the various agencies of the legislative and judicial system. Preliminary investigations provide encouraging evidence that the the IWD can indeed be developed as a viable countermeasure option.

Further research is required in the following areas:

a) optimization of task software (scoring strategies, etc.) for in-vehicle use;
b) optimization of IWD physical design for in-vehicle use;
c) determination of IWD sensitivity to fatigue, drugs and other stressors; and
d) demonstration of IWD potential as a drunk driving countermeasure for convicted offenders.

ACKNOWLEDGEMENTS

The authors wishes to acknowledge the efforts of Mr. Mark Anderson and Ms. Teresa Salter who have assisted in the conduct of the experiment and the analysis of the data. The contributions of Dr. Richard Frecker in the development of a pharmacologically sound protocol, in conducting the physical examinations, and in providing interpretations of pharmacological data are gratefully acknowledged. The authors thank NHTSA for providing a vehicle equipped with the DDWS for evaluation of the CTT. Finally, the authors thank Dr. Leslie Buck of the National Research Council for making the Tracometer available for the study and for his assistance in the analysis of the data.

REFERENCES

ALLEN, R.W., STEIN, A.C., SUMMERS, L.G., and COOK, M.L., Drunk Driving Warning System (DDWS)., Volume I, System Concept and Description, Final Report, November, 1983.

ALLEN, R.W., STEIN, A.C., SUMMERS, L.G., and COOK, M.L., Drunk Driving Warning System (DDWS), Volume II, Field Test Evaluation, Final Report December 1983.

BEIRNESS, D., VOGEL-SPROTT, M., and SPROUL, K., Alcohol Effect on Psychomotor Performance as a Function of Prior Practice. (In preparation, 1980)

BEIRNESS D.J., HAAS, G.C., WALSH, P.J., and DONELSON, A.C., Alcohol and Fatal Road Accidents in Canada: A Statistical Look at its Magnitude and Persistence, TIRF, Impaired Driving Report No. 3, January, 1985.

BUCK, L., Skilled Performance on the NRC Stressalyser, National Research Council Reports LTR-CS-130 and 137, 1974.

BUCK, L., LEONARDO, R., and HYDE, F., Measuring Impaired Performance with the NRC Stressalyser, Applied Ergonomics, 12, 231-236, 1981.

BUCK, L., HYDE, F., ISNOR, C.D., LEONARDO, R., and TRUMBLEY, K., Operating Procedures for the NRC Stressalyser, National Research Council Report, LTR-CS-230, June, 1982.

BUCK, L., Performance Measurement by Means of the NRC Tracometer, Human Factors Association of Canada, 16th Annual Meeting Proceedings, Hamilton, Ontario, Training and Performance Assessment, pp 22-25, October 29, 1983.

CAHALAN, D., CISIN, I.H., and CROSSLEY, H.M., American Drinking Practices. A National Study of Drinking Behaviour and Attitudes, New Haven, College and University Press, 1969.

NOY, Y.I., and SMILEY, A., Comparative Effects of Single vs. Multiple Dosing of Marijuana on Psychomotor Tracking Performance, DCIEM No. 84-R-09, March 1984.

NOY, Y.I., A Comparison of the Sensitivity of the Tracometer and the Critical Tracking Task to Alcohol Intoxication, DCIEM No. 85-R-29, July 1985.

SMART, R.G., GLASSER F.B. et al., Research Advances in Alcohol and Drug Problems, Vol. 7, (see Classical Conditioning: Drug Tolerance and Drug Dependence), 1983.

SNYDER, M.B., The Drunk Driving Warning System - Status Review, Presented at the 64rd Annual Meeting Transportation Research Board, Washington, D.C., January 18, 1984.

ANNEX A

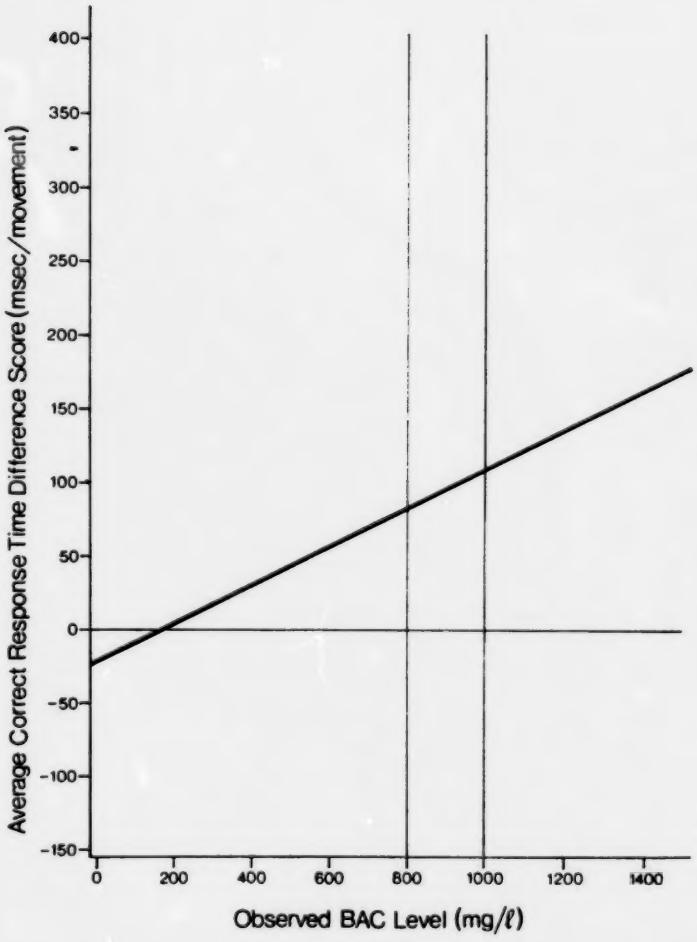
DOSE-RESPONSE CURVES FOR SEVEN TRACOMETER VARIABLES

Listing of Figures

A1 - Correct Response Time (CRT)
A2 - Non-Overshoot Movement Time (NOM)
A3 - Overshoot Movement Time (OM)
A4 - Overshoot Score (OS)
A5 - Error Score (ES)
A6 - Total Response Time (TRT)
A7 - Reaction Time (RT)

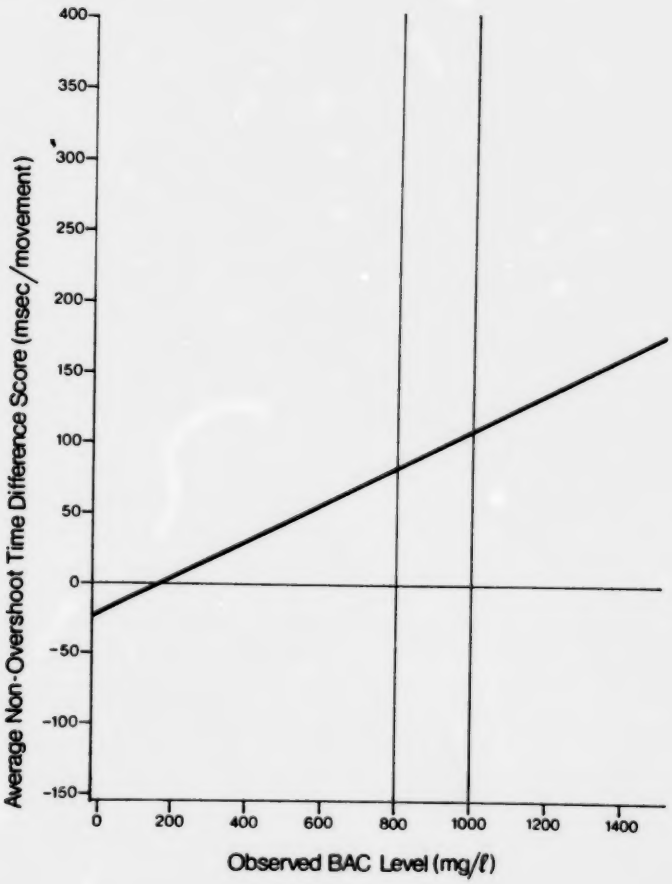
A1

Correct Response Time



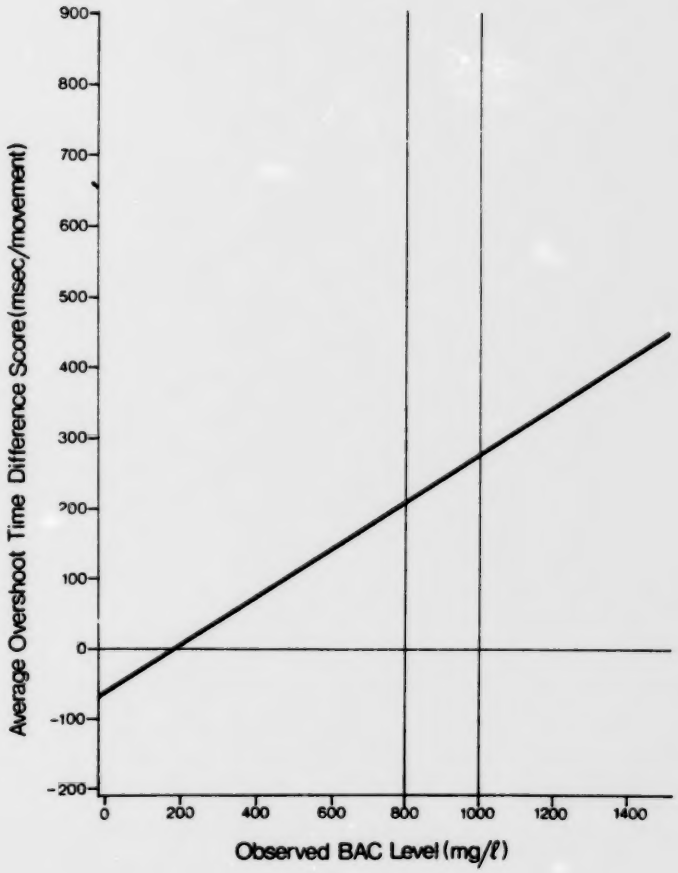
A2

Non-Overshoot Movement Time



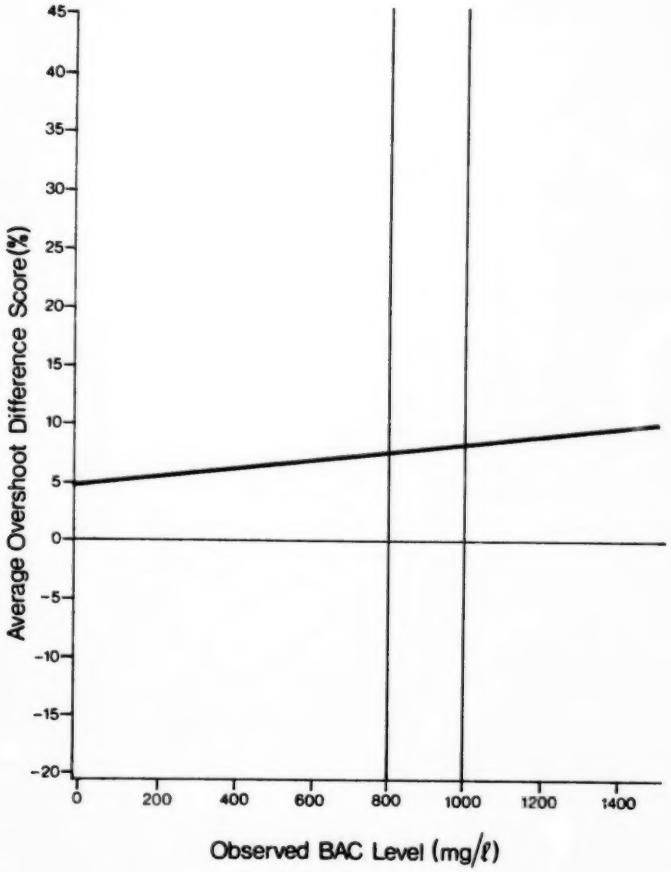
A3

Overshoot Movement Time

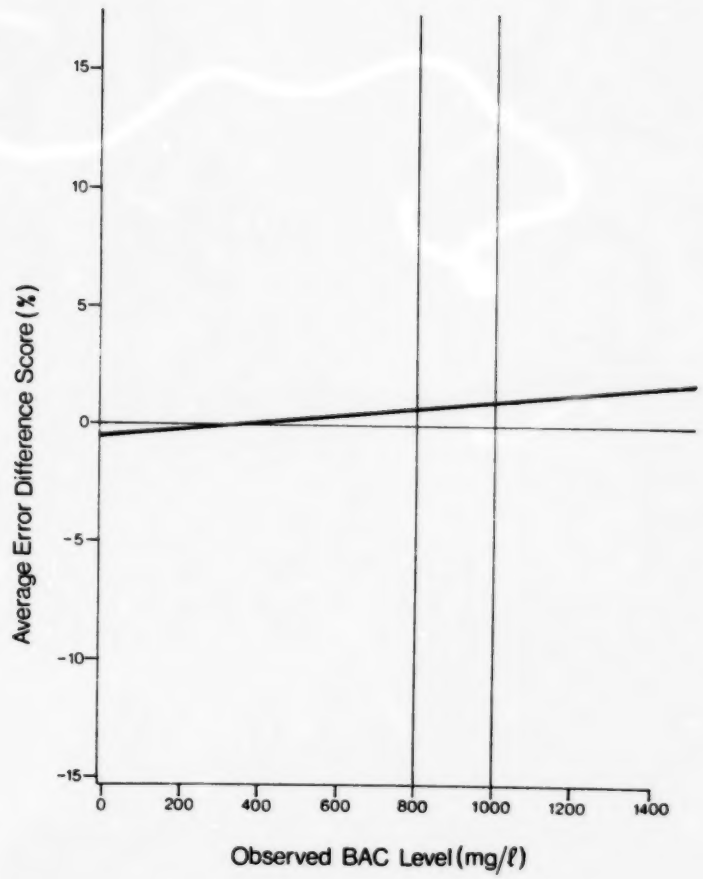


A4

Overshoot Score

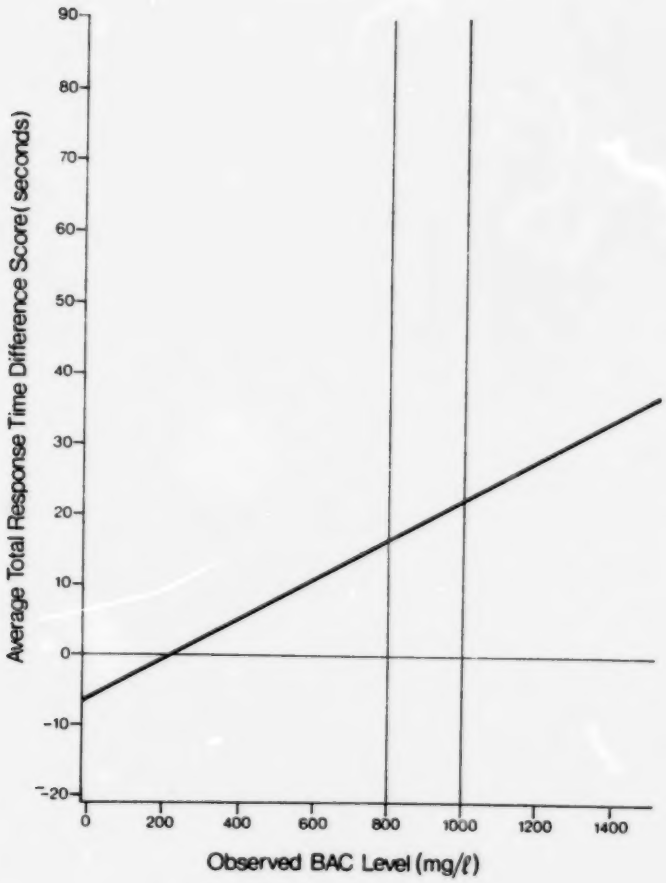


A5
Error Score



A6

Total Response Time



A7

Response Time Scores and Total Response Time Scores